Large deviation results for $U$- and $V$-statistics
and $U$- and $V$-empiricals

Robert Serfling$^1$
University of Texas at Dallas

and

Wenyang Wang$^2$
Crestar Bank

December, 1998

$^1$Programs in Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Supported by NSF Grant DMS-9705209.

$^2$Crestar Bank, Richmond, Virginia, USA
Abstract

This paper develops large deviation results for empirical df’s of $U$-statistic structure (i.e., $U$-empirical df’s), for associated “von Mises” analogues (i.e., $V$-empirical df’s), and for statistics defined as functionals evaluated at these entities. The results extend the work of Groeneboom, Oosterhoff, and Ruymgaart (1979) on classical empirical df’s and functionals. Along the way we also extend to a topology stronger than the $\tau$ topology the basic Sanov theorem for the classical empirical df and we obtain some improvements of LD results of Eichelsbacher and Löwe (1995) for $U$- and $V$-statistics. Applications treated include “generalized $L$-statistics,” with special attention to trimmed $U$-statistics, certain nonparametric spread measures, and goodness-of-fit statistics of weighted Kolmogorov-Smirnov type.

AMS 1991 subject classifications. 60F10, 62E20.

Key words and phrases. Large deviations, $U$-statistics, $U$-empirical measures, Sanov theorem, generalized $L$-statistics, trimmed $U$-statistics, nonparametric spread measures, goodness-of-fit statistics, weighted Kolmogorov-Smirnov statistics, $\tau$ topology.
1 Introduction

The classical (weak) law of large numbers asserts, for the usual sample mean of i.i.d. random variables having mean $\mu$, that $P(|\bar{X}_n - \mu| > \epsilon) \to 0$ as $n \to \infty$ for every $\epsilon > 0$. “Large deviation theory” treats the rate of convergence of these probabilities to 0 and, in particular, characterizes the exponential index that applies under typical conditions. Initiated by Cramér (1938), the theory has enjoyed rich development and has found diverse applications including asymptotic relative efficiencies of statistical tests of hypotheses, sequential fixed-width confidence interval procedures, optimality of estimation procedures, fast simulation methods for analysis of communication systems, detection theory, analysis of diffusion processes, statistical mechanics, occupation times for Markov chains and processes, and perturbations of random dynamical systems. For background, see [7], [33], [1], [2], [21], [34], [10], [15], [5], [9], [26, Chap. 10], and [29].

Analogous to the treatment of the sample mean, large deviation theory of the sample distribution function has also received considerable investigation, initially by Sanov (1957), and later by Hoadley (1967) and Groeneboom, Oosterhoff, and Ruymgaart (1979), among others. I.e., for the classical empirical measure of $X_1, \ldots, X_n$,

$$L_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i},$$

where $\delta_x$ denotes the Dirac measure with unit mass at $x$, the exponential rate of convergence to 0 of probabilities of the form $P(L_n \in A)$ is established for suitable choices of set $A$ of df’s excluding the cdf $F$ of the observations. For statistics defined as “statistical functionals” $T(L_n)$, the corresponding LD theory is then obtained by taking sets $A$ of the form $\{G : T(G) \geq t\}$. A very general version of “Sanov’s theorem” has been given by de Acosta (1994).

The main purposes of the present paper are to develop LD results for empirical df’s of $U$-statistic structure (i.e., $U$-empirical df’s), for associated “von Mises” analogues (i.e., $V$-empirical df’s), and for statistics defined as functionals evaluated at these entities. The studies of $U$- and $V$-statistics (i.e., averages of the evaluations of a kernel $h(x_1, \ldots, x_m)$ over sets of observations taken $m$ at a time) were initiated by Hoeffding (1948) and von Mises (1947), respectively, and serve a wide scope of statistical applications (for background, see [26, Chaps. 5 & 6] and [22]). The investigation of $U$-statistics and related entities has, for the most part, centered around weak convergence and almost sure behavior, until recently the topic of large deviation theory has received serious attention (see, for example, Eichelsbacher and Löwe (1995)). Here we extend to $U$- and $V$-empiricals and to functionals defined on these, and along the way we obtain an extension to a stronger topology of the basic Sanov theorem for $L_n$. For earlier versions of our results, see Wang (1994).

We now introduce the setting of the paper and describe precisely the main contributions. The following standard notation will be used throughout. For any pointwise function $g(x)$ defined on a set $\mathcal{X}$, we define an associated set function on the subsets of $\mathcal{X}$ by $g(A) = \inf_{x \in A} g(x)$. For any measure $\mu$ and $\mu$-integrable function $f$, we put $\mu(f) = \int f d\mu$. For any set $A$, we denote by $\text{int}(A)$ and $\text{cl}(A)$ the interior and closure of $A$, respectively.

By large deviation principle (LDP), we mean the following. (See, e.g., Dembo and Zeitouni (1993), §1.2, or Varadhan (1984), §2.)

Definition 1.1 Let $\mathcal{Y}$ be a topological space and $\mathcal{A}$ a $\sigma$-field in $\mathcal{Y}$ (not necessarily the Borel $\sigma$-field). A family of probability measures $\{\mu_\epsilon\}$ on $\mathcal{A}$ satisfies the large deviation principle (LDP)
with rate function $I$, $0 \leq I(y) \leq \infty$ for $y \in \mathcal{Y}$, if the level sets $\{y \in \mathcal{Y} : I(y) \leq c\}$ for all $c < \infty$ are compact subsets of $\mathcal{Y}$, and for each $A \in \mathcal{A}$,

$$-I(\text{int}(A)) \leq \liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(A) \leq \limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(A) \leq -I(\text{cl}(A)).$$

(1)

A sequence of random elements of $\mathcal{Y}$ satisfies the LDP if the corresponding probability measures do.

The broadest statistical applications of an LDP are obtained when the relevant topology is as strong as possible. Thus much of the work of LD theory is devoted to establishing the LD rate in the strongest possible topological setting.

Let $(S, d)$ be a separable metric space with Borel $\sigma$-field $\mathcal{S}$, and, for integer $m \geq 1$, let $S^m$ be the $m$-fold product measurable space with Borel $\sigma$-field $\mathcal{S}^m = \mathcal{S} \times \cdots \times \mathcal{S}$. Let $B(S^m)$ denote the set of bounded measurable functions on $S^m$, $B_c(S^m)$ the set of continuous functions in $B(S^m)$, and $\mathcal{P}(\mathcal{S}^m)$ the set of probability measures on $\mathcal{S}^m$. The weak (resp., $\tau$) topology on $\mathcal{P}(\mathcal{S}^m)$ is the smallest topology such that for each function $f$ in $B_c(S^m)$ (resp., $B(S^m)$), the map $\mu \mapsto \mu(f)$, $\mu \in \mathcal{P}(\mathcal{S}^m)$, is continuous. Further, we define $B_w$ (resp., $B_\tau$) to be the smallest $\sigma$-field in $\mathcal{P}(\mathcal{S}^m)$ such that for each function $f$ in $B_c(S^m)$ (resp., in $B(S^m)$), the map $\mu \mapsto \mu(f)$, $\mu \in \mathcal{P}(\mathcal{S}^m)$, is measurable. Let $(X_i, i \geq 1)$ be a sequence of independent $S$-valued random variables with common law $P \in \mathcal{P}(\mathcal{S})$. For a “kernel” $h$ mapping $(S^m, \mathcal{S}^m)$ to an arbitrary measurable space $(T, T)$, with summation defined on $T$, and for sample size $n \geq m$, we denote the corresponding $U$-statistic by

$$U_n(h) = n^{-\frac{1}{m}} \sum_{1 \leq i_1 \neq \cdots \neq i_m \leq n} h(X_{i_1}, \ldots, X_{i_m})$$

and the corresponding $V$-statistic by

$$V_n(h) = n^{-m} \sum_{i_1 = 1}^{n} \cdots \sum_{i_m = 1}^{n} h(X_{i_1}, \ldots, X_{i_m}).$$

(Typically, $T$ will be some Euclidean space $\mathbb{R}^p$, usually just $\mathbb{R}$.)

For estimation of the probability measure $P^m$ of $(X_1, \ldots, X_m)$, the $U$- and $V$-empirical measures are, respectively, the random elements of $\mathcal{P}(\mathcal{S}^m)$ given by

$$\hat{G}_n = n^{-\frac{1}{m}} \sum_{1 \leq i_1 \neq \cdots \neq i_m \leq n} \delta_{(X_{i_1}, \ldots, X_{i_m})}$$

and

$$\tilde{G}_n = n^{-m} \sum_{i_1 = 1}^{n} \cdots \sum_{i_m = 1}^{n} \delta_{(X_{i_1}, \ldots, X_{i_m})} = L_n^m.$$

Note that $U_n(h) = \int_{S^m} h \, d\hat{G}_n = \hat{G}_n(h)$ and $V_n(h) = \int_{S^m} h \, d\tilde{G}_n = \tilde{G}_n(h)$.

For estimation of the probability measure $P^m \circ h^{-1}$ of $h(X_1, \ldots, X_m)$, the kernel-type $U$- and $V$-empirical measures are the random elements of $\mathcal{P}(T)$ given, respectively, by

$$\hat{H}_n = \hat{G}_n \circ h^{-1} = n^{-\frac{1}{m}} \sum_{1 \leq i_1 \neq \cdots \neq i_m \leq n} \delta_{h(X_{i_1}, \ldots, X_{i_m})}$$

and

$$\tilde{H}_n = \tilde{G}_n \circ h^{-1} = n^{-m} \sum_{i_1 = 1}^{n} \cdots \sum_{i_m = 1}^{n} \delta_{h(X_{i_1}, \ldots, X_{i_m})}.$$
and
\[ \tilde{H}_n = \tilde{G}_n \circ h^{-1} = n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \delta_h(x_{i_1},\ldots,x_{i_m}). \]

In Section 3 we develop (Theorem 3.1) LDP’s in the \( \tau \) topology for the above four \( U \)- and \( V \)-empiricals. Further, in Corollary 3.1 we obtain the LDP for the classical empirical df \( L_n \) in a topology somewhat stronger than the \( \tau \) topology of previous results [16], [8], and in Conjecture 3.1 we formulate a possible further strengthening. In Theorem 3.2 we provide the LDP for suitable (i.e., \( \tau \)-continuous) functionals of \( U \)- and \( V \)-empiricals. These results extend the work of [16] on classical empiricals. In Example 3.1 we make application to “generalized \( L \)-statistics,” with special attention to trimmed \( U \)-statistics and certain nonparametric spread measures. In Example 3.2 we treat the LDP for members of a class of goodness-of-fit statistics of weighted Kolmogorov-Smirnov type. A basic tool used in Section 3 consists of a preliminary lemma (Lemma 3.1) on the “projective limit” approach.

In Section 2 we provide various tools and foundational results, including the version of Sanov’s theorem of [8] (a tool that we shall use, although a slight improvement is obtained in Section 3), and several versions of “contraction principle” (a standard tool in LD analysis). Also, we derive versions of the LDP for \( U \)- and \( V \)-statistics (Theorem 2.1, Corollary 2.1, Theorem 2.2) which are similar to those of [14] but incorporate some improved methods of proof and statements of conditions.

2 Preliminary results, including LDP’s for \( U \)- and \( V \)-statistics

2.1 Preliminary lemmas

For two probability measures \( P \) and \( Q \), the well-known Kullback-Leibler information number (or relative entropy) of \( Q \) with respect to \( P \) is defined as
\[ K(Q,P) = \begin{cases} \int (dQ/dP) \log(dQ/dP) \, dP & \text{if } Q \ll P \\ \infty & \text{otherwise.} \end{cases} \]

We will make use of the following lemma based on Theorem 1.1 and Lemma 2.1 of de Acosta (1994).

**Lemma 2.1** Let \((S,\mathcal{S})\) be an arbitrary measurable space and let the i.i.d. sequence \(\{X_i\}\) have probability measure \(P\). Then \(L_n\) satisfies the LDP on \(\mathcal{P}(S)\) in the \(\tau\) topology (i.e., considered as a \(\mathcal{B}_\tau\)-measurable function) with rate function \(J_P(Q) = K(Q,P)\), \(Q \in \mathcal{P}(S)\).

**Remarks 2.1** (i) This result extends previous work of Sanov (1957), Hoadley (1967), and Groeneboom, Oosterhoff, and Ruymgaart (1979) and appears to represent the most general version of what is known as Sanov’s theorem for i.i.d. sequences.

(ii) In Section 3 (see Corollary 3.1), however, we strengthen this result to a stronger topology.

(iii) For our needs below, however, we will use only the (immediate) implication of Lemma 2.1 that \(L_n\) satisfies the LDP on \(\mathcal{P}(S)\) in the weak topology. \(\square\)

Next we state several so-called “contraction principles”, which provide very useful tools for deriving LDP’s under continuous and certain “approximately” continuous transformations of measures already satisfying an LDP. We refer to [9, Chap. 4] for more detailed discussion.
Lemma 2.2 (a) Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces and $f : \mathcal{X} \rightarrow \mathcal{Y}$ a continuous function. If a family of probability measures $\{\mu_n\}$ on $\mathcal{X}$ satisfies the LDP with rate function $I$, then the family of probability measures $\{\mu_n \circ f^{-1}\}$ satisfies the LDP on $\mathcal{Y}$ with rate function $I_f(y) = \inf\{I(x) : x \in \mathcal{X}, y = f(x)\}$. (See, e.g., [9, Thm. 4.2.1].)

(b) Let $\{\mu_n\}$ be a family of probability measures on a complete separable metric space $\mathcal{X}$ satisfying the LDP with rate function $I$. Let $f_k$ be continuous maps from $\mathcal{X} \rightarrow \mathcal{Y}$, where $\mathcal{Y}$ is another complete separable metric space. Assume that $\lim_{n \rightarrow 0} f_k = f$ exists uniformly over compact subsets of $\mathcal{X}$. Then $\{\mu_n \circ f_k^{-1}\}$ satisfies the LDP with rate function $I_f(y) = I(f^{-1}(y))$. (See, e.g., [33, Thm. 2.4].)

(c) Let $\{\mu_n\}$ be a family of probability measures that satisfies the LDP with rate function $I$ on a Hausdorff topological space $\mathcal{X}$, and for $k \geq 1$, let $f_k : \mathcal{X} \rightarrow \mathcal{Y}$ be continuous mappings, with $(\mathcal{Y}, d)$ a metric space. Let $f : \mathcal{X} \rightarrow \mathcal{Y}$ be a measurable map. If the conditions

(i) $\limsup_{k \rightarrow \infty} \sup_{\{x : f_k(x) \leq \alpha\}} d(f_k(x), f(x)) = 0$ for every $\alpha < \infty,$

(ii) $\lim_{n \rightarrow \infty} \limsup_{\epsilon \rightarrow 0} \epsilon \log(\mu_n \{x \in \mathcal{X} : d(f_k(x), f(x)) \geq \delta\}) = -\infty$ for every $\delta > 0,$

both hold, then $\{\mu_n \circ f_k^{-1}\}$ satisfies the LDP on $\mathcal{Y}$ with rate function $I_f(y) = I(f^{-1}(y))$. (See, e.g., [9, Thm. 4.2.23].)

The next result gives LDP’s for the $U$- and $V$-empiricals $\tilde{G}_n$ and $\hat{G}_n$ in the weak topology on $\mathcal{P}(\mathcal{S}^{(m)})$.

Lemma 2.3 Let $(S, d)$ be a separable metric space with Borel $\sigma$-field $\mathcal{S}$ and let $\{X_i\}$ be an i.i.d. $S$-valued sequence having probability measure $P$. Then the $U$- and $V$-empiricals $\tilde{G}_n$ and $\hat{G}_n$ each satisfy the LDP on $\mathcal{P}(\mathcal{S}^{(m)})$ in the weak topology (i.e., considered as $\mathcal{B}_w$-measurable functions) with rate function

$$J_P^{(m)}(\mu) = \begin{cases} K(Q, P) & \text{if } \mu = Q^m \in \mathcal{P}(\mathcal{S}^{(m)}) \\ \infty & \text{otherwise.} \end{cases}$$

Proof. We first establish the result for the $V$-empirical $\tilde{G}_n = L_n^m$. Since the mapping $f : \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{P}(\mathcal{S}^{(m)})$ defined by $f(Q) = Q^m$ is continuous with respect to the respective weak topologies, we apply Lemma 2.1 and Lemma 2.2(a) to obtain that $\tilde{G}_n$ satisfies the LDP on $\mathcal{P}(\mathcal{S}^{(m)})$ in the weak topology with rate function $J_P^{(m)}(\mu) = J_P(f^{-1}(\mu)) = J_P(Q) = K(Q, P)$ if $\mu = Q^m$, and $= \infty$ otherwise, for $\mu \in \mathcal{P}(\mathcal{S}^{(m)})$.

Next we show that $\tilde{G}_n$ must satisfy the same LDP. For this we use Proposition 1 of Baxter and Jain (1988), which gives the following criterion for two sequences of probability measures $\{\mu_n\}$ and $\{\nu_n\}$ on the Borel $\sigma$-field in a metric space $(\mathcal{X}, d)$ to have the same LDP behavior: for every closed set $A$ of $\mathcal{X}$ and every $\epsilon > 0$,

$$\limsup_{n \rightarrow \infty} (1/n) \log \mu_n(A) \leq \limsup_{n \rightarrow \infty} (1/n) \log \nu_n(A^c) \quad (2)$$

$$\limsup_{n \rightarrow \infty} (1/n) \log \nu_n(A) \leq \limsup_{n \rightarrow \infty} (1/n) \log \mu_n(A^c), \quad (3)$$

where $A^c = \{y \in \mathcal{X} : d(y, A) < \epsilon\}$. To apply this criterion, we let $\{\mu_n\}$ and $\{\nu_n\}$ denote the probability measures of $\tilde{G}_n$ and $\hat{G}_n$, respectively, and metrize the weak topology on $\mathcal{P}(\mathcal{S}^{(m)})$ by the Prohorov metric

$$\rho(P, Q) = \inf\{\epsilon > 0 : P(F) \leq Q(F^c) + \epsilon, \text{ all closed sets } F \subset S^m\}, \quad P, Q \in \mathcal{P}(\mathcal{S}^{(m)}),$$

4
where \( F^\varepsilon = \{ y \in S^m : d_{(m)}(y, F) < \varepsilon \} \) and \( d_{(m)} \) is a metric on \( S^m \) induced by the metric \( d \). It is immediate that for any closed set \( A \subseteq \mathcal{P}(\mathcal{S}^{(m)}) \)

\[
\{ \tilde{G}_n \in A \} \subseteq \{ \tilde{G}_n \in A^\varepsilon \} \cup \{ \rho(\tilde{G}_n, \tilde{G}_n) \geq \varepsilon \},
\]

whence

\[
\mu_n(A) \leq \nu_n(A^\varepsilon) + P(\rho(\tilde{G}_n, \tilde{G}_n) \geq \varepsilon).
\]

Similarly,

\[
\nu_n(A) \leq \mu_n(A^\varepsilon) + P(\rho(\tilde{G}_n, \tilde{G}_n) \geq \varepsilon).
\]

Now

\[
\rho(\tilde{G}_n, \tilde{G}_n) \leq \sup_{B \in \mathcal{S}^{(m)}} |\tilde{G}_n(B) - \tilde{G}_n(B)| \leq 2 \left( \frac{n^m - n^{(m)}}{n^m} \right) = O(n^{-1})
\]

\[
< \varepsilon \text{ for } n \text{ sufficiently large.}
\]

Hence \( P(\rho(\tilde{G}_n, \tilde{G}_n) \geq \varepsilon) = 0 \) for all \( n \) sufficiently large and the Baxter/Jain criterion is satisfied. \( \square \)

### 2.2 LDP’s for \( U^- \) and \( V \)-statistics

Via Lemmas 2.2 and 2.3 we obtain LDP’s for \( U^- \) and \( V \)-statistics. Our approach follows that of Eichelsbacher and Löwe (1995), with certain modifications. In particular, by treating \( U^- \) and \( V \)-statistics separately, we are able to relax somewhat the moment generating function conditions on the kernel in the \( U \)-statistic case.

**Theorem 2.1** Let \( h \) be an \( \mathbb{R}^p \)-valued kernel defined on \( S^m \). Assume that \( E \exp(t\|h\|) < \infty \) for all \( t > 0 \), where \( \| \cdot \| \) denotes the usual Euclidean norm. Then \( U_n(h) \) satisfies the LDP on \( \mathbb{R}^p \) in the usual topology with rate function \( I_h(x) = \inf\{K(Q, P) : \int_{S^m} h dQ^m = x, Q^m \in \mathcal{P}(\mathcal{S}^{(m)})\} \), \( x \in \mathbb{R}^p \).

**Proof.** Recall that \( U_n(h) = f(\tilde{G}_n) \), where \( f \) is the linear functional from \( \mathcal{P}(\mathcal{S}^{(m)}) \) to \( \mathbb{R}^p \), \( f(\mu) = \int_{S^m} h d\mu \). If \( h \) is bounded and continuous, then the functional \( f \) is continuous with respect to the weak topology on \( \mathcal{P}(\mathcal{S}^{(m)}) \) and the claim follows immediately from Lemma 2.3 with the use of Lemma 2.2(a).

For an arbitrary bounded kernel \( h \), \( \|h\| \leq C \) for some finite \( C > 0 \), it is a straightforward consequence of Lusin’s theorem that there exists a sequence of continuous functions \( \{h_k, k = 1, 2, \ldots\} \) satisfying \( \|h_k\| \leq C \) for each \( k \), and \( P^m(\lim_{k \to \infty} h_k = h) = 1 \). (For proof see Rudin (1987), p. 56, and for proof of Lusin’s theorem for normal topological spaces, which include metric spaces, see Dudley (1989), p. 190). In view of Lemma 2.2(c), we need to verify the following two conditions:

\[
\lim_{k \to \infty} \sup_{\alpha < \infty} \left\{ \left\| \int_{S^m} (h_k - h) dQ^m \right\| : K(Q, P) \leq \alpha \right\} = 0,
\]

and

\[
\lim_{k \to \infty} \limsup_{n \to \infty} n^{-1} \log P(\|U_n(h_k - h)\| \geq \delta) = -\infty, \text{ each } \delta > 0.
\]

Condition (4) follows from the fact that \( \lim_{k \to \infty} \| \int_{S^m} (h_k - h) dP^m \| = 0 \) and the property that the collection \( \{ Q \in \mathcal{P}(\mathcal{S}) : K(Q, P) \leq \alpha \} \) is uniformly absolutely continuous with respect to \( P \). For (5) it suffices to consider a real-valued kernel \( h \) and show that

\[
\lim_{k \to \infty} \limsup_{n \to \infty} n^{-1} \log P(U_n(h_k - h) \geq \delta) = -\infty.
\]
By an exponential probability inequality for $U$-statistics (cf. Serfling (1980), §5.6), and for $t > 0$, 
\[ P(U_n(h_k - h) \geq \delta) \leq \exp(-\delta[n/m]t) (E \exp(t(h_k - h)))^{[n/m]}, \]
where $[n/m]$ denotes the integer part of $n/m$, we have
\[
\lim_{n \to \infty} n^{-1} \log P(U_n(h_k - h) \geq \delta) \leq m^{-1}(-\delta t + \log E \exp(t(h_k - h))),
\]
which yields (5) by the bounded convergence theorem. Extension to $\mathbb{R}^p$-valued kernels is straightforward. Thus the claim holds for $U$-statistics with bounded kernels.

For the general case of an arbitrary kernel $h$, let $h_k(x) = h(x)$ if $\|h(x)\| \leq k$, and $h_k(x) = 0$ otherwise, $k \geq 1$. Then $\lim_{k \to \infty} h_k(x) = h(x)$ for all $x \in S^m$, and $\|h_k - h\| \leq \|h\|$ for each $k \geq 1$. Condition (4) holds by the same token as before, and condition (5) follows from the assumed moment generating function condition and the dominated convergence theorem.

**Remark 2.2** The above LDP is obtained by Eichelsbacher and Löwe (1995) in a single result formulated for both $U$- and $V$-statistics that imposes certain additional moment generating function restrictions. The additional restrictions, however, are needed only for the $V$-statistic case, which we treat here in Corollary 2.1 below.

In stating the assumptions for the LDP for $V$-statistics, we adopt the notation of [14]. For $1 \leq r \leq m$, let $\mathcal{H}_r$ be the set of all versions of $h$ depending on exactly $r$ variables only. That is, $g \in \mathcal{H}_r$ if and only if for some (ordered) partition $(P^1, \ldots, P^r)$ of $\{1, \ldots, m\}$ we have $g(y_1, \ldots, y_r) = h(x_1, \ldots, x_m)$ for some choice of $(y_1, \ldots, y_r)$ such that, for $1 \leq j \leq r$, $y_j = x_i$ for some $i \in P^j$.

**Corollary 2.1** Assume the conditions of Theorem 2.1. Further, assume for each $r = 1, \ldots, m - 1$ and for every $g \in \mathcal{H}_r$, that $E \exp(t_0 \|g\|) < \infty$ for some constant $t_0 > 0$. Then $V_n(h)$ satisfies the LDP with the same rate function $I_h$ defined in Theorem 2.1.

**Proof.** By an argument similar to our use of [3, Prop. 1] in proving Lemma 2.3, it suffices to show that for each $\delta > 0$,
\[
\lim_{n \to \infty} n^{-1} \log P(\|V_n(h) - U_n(h)\| \geq \delta) = -\infty. \tag{6}
\]
Since 
\[ V_n(h) = \frac{n(m)}{n^m} U_n(h) + \frac{1}{n^m} \sum_{1 \leq r \leq m-1, g \in \mathcal{H}_r} n(r) U_n(g), \]
we have
\[
P(\|V_n(h) - U_n(h)\| \geq \delta)
\leq P \left( \|U_n(h)\| \geq \frac{n^m}{n^m - n(m)} \frac{\delta}{2} \right) + \sum_{r \geq 1} n(r) P \left( \left\| \frac{1}{n^m} \sum_{1 \leq r \leq m-1, g \in \mathcal{H}_r} n(r) U_n(g) \right\| \geq \frac{\delta}{2} \right)
\leq P \left( \|U_n(h)\| \geq \frac{n^m}{n^m - n(m)} \frac{\delta}{2} \right) + \sum_{r \geq 1} n(r) M \left( \left\| U_n(g) \right\| \geq \frac{n^m}{n(r)} \frac{\delta}{2M} \right),
\]
where $M$ denotes the cardinality of the set $\bigcup_{r=1}^{m-1} \mathcal{H}_r$. We now obtain (6) by applying to each term the exponential probability inequality for $U$-statistics as in the proof of Theorem 2.1.
We next establish, for the case of polynomial kernels, LDP’s for $U$- and $V$-statistics under less stringent mgf conditions than those imposed by Theorem 2.1 and Corollary 2.1. The following preliminary lemma treats random variables which can be expressed as functions of sample moments, e.g., sample central moments and Fisher’s $k$-statistics. For these the LDP follows by straightforward application of the contraction principle and Cramér’s large deviation theorem for sample means in the multi-dimensional setting. Since the Cramér LDP has been established under quite relaxed mgf conditions (see [9] for a detailed treatment), this leads to LDP’s for $U$- and $V$-statistics under similarly less stringent mgf conditions. Our treatment of $U$- and $V$-statistics for the special case of polynomial kernels is detailed more fully in [36]. A similar treatment is given by [14], where, however, more stringent mgf conditions are adopted.

**Lemma 2.4** Let $X_1, X_2, \ldots$ be i.i.d. real-valued random variables. For $k \geq 1$, define $\hat{\mu}_{kn} = n^{-1} \sum_{i=1}^{n} X_i^k$. For $p$ be a fixed positive integer, let $f_n : \mathbb{R}^p \to \mathbb{R}$ be continuous functions such that $f_n \to f$ as $n \to \infty$ uniformly over compact sets in $\mathbb{R}^p$. Assume that $M(t) = E \exp(\sum_{i=1}^{p} t_i X_i^k)$ is finite for all $t = (t_1, \ldots, t_p)$ in a neighborhood of the origin in $\mathbb{R}^p$. Then $\{f_n(\hat{\mu}_{1n}, \ldots, \hat{\mu}_{pn})\}$ satisfies the LDP with rate function $I(f^{-1}(x))$, where $I(y) = \sup_{t \in \mathbb{R}^p} \{ \langle t, y \rangle - \log M(t) \}$ and $\langle t, y \rangle$ denotes the inner product of $t, y \in \mathbb{R}^p$.

**Proof.** By Cramér’s LDP for the mean of i.i.d. variables in $\mathbb{R}^p$ (see, e.g., [33, Thm. 4.1] and [9, Cor. 6.1.6]), the sequence $\{(\hat{\mu}_{1n}, \ldots, \hat{\mu}_{pn})\}$ satisfies the LDP in $\mathbb{R}^p$ with the rate function $I(y)$ defined above. Now apply Lemma 2.2(b).

**Theorem 2.2** Let $h$ be a kernel of polynomial form,

$$h(x_1, \ldots, x_m) = \sum_{(r_1, \ldots, r_m) \in \Xi} c_{r_1 \ldots r_m} x_1^{r_1} \cdots x_m^{r_m},$$

where $\Xi$ is a finite index set consisting of selected $m$-tuples of nonnegative integers, and $c_{r_1 \ldots r_m}$ are constants. Let $r = \max\{r_1, \ldots, r_m : (r_1, \ldots, r_m) \in \Xi\}$, and $s = \max\{r_1 + \cdots + r_m : (r_1, \ldots, r_m) \in \Xi\}$. Assume that $E \exp(t_0 |X|^r) < \infty$ for some $t_0 > 0$. Then the $V$-statistics, $V_n(h)$, satisfy the LDP with rate function $I(\hat{h}^{-1})$, where $I$ is as in Lemma 2.4 and $\hat{h}$ is given by

$$\hat{h}(x_1, \ldots, x_r) = \sum_{(r_1, \ldots, r_m) \in \Xi} c_{r_1 \ldots r_m} x_1^{r_1} \cdots x_m^{r_m},$$

with $x_0 = 1$. Assume, further, that $E \exp(s_0 |X|^s) < \infty$ for some $s_0 > 0$. Then the $U$-statistics, $U_n(h)$, satisfy the same LDP.

**Proof.** We first consider the $V$-statistics $V_n(h)$,

$$V_n(h) = n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h(X_{i_1}, \ldots, X_{i_m})$$

$$= n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \sum_{(r_1, \ldots, r_m) \in \Xi} c_{r_1 \ldots r_m} X_{i_1}^{r_1} \cdots X_{i_m}^{r_m}$$

$$= \sum_{(r_1, \ldots, r_m) \in \Xi} c_{r_1 \ldots r_m} \hat{\mu}_{r_1} \cdots \hat{\mu}_{r_m}$$

$$= \hat{h}(\hat{\mu}_{1n}, \ldots, \hat{\mu}_{rn}).$$
Hence, by Lemma 2.4, $V_n(h)$ satisfies the LDP with rate function $I(\hat{\mu}^{-1})$. For the LDP for U-statistics, observe that the moment generating function condition $E \exp(s_0|X|^r) < \infty$, for some $s_0 > 0$, implies via Hölder’s inequality that for some constant $t > 0$, $E \exp(|t|g|) < \infty$ for all $g \in H_r$ and $1 \leq r \leq m$. Thus the proof of Corollary 2.1 can be adapted with minor modification.

Remarks 2.3 (i) The moment generating function condition of Lemma 2.4 is equivalent to the condition that $E \exp(t_0|X|^p) < \infty$ for some $t_0 > 0$. Moreover, such a condition cannot be completely omitted (see Slaby (1988) and Dinwoodie (1991) for examples of the failure of the LDP when such a condition is not assumed).

(ii) The mgf condition imposed by Corollary 2.1 for the LDP for V-statistics is stronger than that imposed by Theorem 2.1 for the LDP for U-statistics. On the other hand, the mgf conditions imposed by Theorem 2.2, in the special case of polynomial kernels, are weaker for the case of V-statistics than for the case of U-statistics. Also, as illustrated in Example 2.1 below, the mgf conditions of Theorem 2.2 are weaker than those imposed by Theorem 2.1 and Corollary 2.1, which, however, treat kernels of arbitrary structure. These disparities are artifacts of the two methods of proof and indicate that definitive conditions for the case of general kernels are not yet known.

Example 2.1 The sample variance. Let $X_1, X_2, \ldots$ be i.i.d. real-valued random variables and consider the kernel

$$h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2.$$

The corresponding U-statistic is the sample variance, which also may be expressed as

$$s_n^2 = (n - 1)^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2,$$

where $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$, and the corresponding V-statistic is $m_n = \frac{n-1}{n} s_n^2$. To apply Theorem 2.2, we introduce the assumption that $E \exp(t_0 X^2) < \infty$ for some $t_0 > 0$ and observe that the kernel $h$ may be represented as

$$h(x_1, x_2) = \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 - x_1 x_2.$$  

This is of the form in Theorem 2.2 with $\Xi = \{(2,0), (0,2), (1,1)\}$ and $c_{(2,0)} = 1/2 = c_{(0,2)}$ and $c_{(1,1)} = -1$. The corresponding $\hat{\mu}$ is then given by

$$\hat{\mu}(x_1, x_2) = x_2 - x_1^2,$$

whence the statistic $s_n^2$ and $m_n$ each satisfy the LDP

$$\lim_{n \to \infty} n^{-1} \log P(s_n^2 > u) = \lim_{n \to \infty} n^{-1} \log P(m_n > u) = - \inf_{(x_1, x_2) \in C_u} I(x_1, x_2),$$

where $C_u = \{(x_1, x_2) : \hat{\mu}(x_1, x_2) > u\}$ for $u > 0$ and

$$I(x_1, x_2) = \sup_{(t_1, t_2) \in \mathbb{R}^2} \{t_1 x_1 + t_2 x_2 - \log E \exp(t_1 X + t_2 X^2)\}.$$  

Here we have used the fact that in the LDP for the sample means, equality in (1) holds for open and convex sets (see, e.g., [2, Thm. 2.3]). Let us compare the above moment condition with the one required for this problem by Theorem 2.1, namely that $E \exp(t(X - Y)^2) < \infty$ for all $t > 0$. This condition is equivalent to $E \exp(t X^2) < \infty$ for all $t > 0$, which is significantly more restrictive than the above condition.
3 LDP’s for $U$- and $V$-empirical measures and related statistical functionals

In this section we treat LDP’s for $U$- and $V$-empirical measures in the $\tau$-topology. Instead of the contraction principle used in the previous section, we shall use the method of “projective limits” (e.g., [9, §4.6]), whereby a collection of LDP’s for a family of spaces is “lifted” into the LDP for the space which is the projective limit of the given spaces. For convenience, we first state and prove the following direct consequence of the projective limit approach.

Lemma 3.1 Let $X$ be an arbitrary set and $Y$ a Hausdorff topological space with the Borel $\sigma$-field $B$. Let $T$ be the topology on $X$ generated by a family of mappings $\Lambda_\alpha : X \to Y$, $\alpha \in A$ (some index set), and the smallest $\sigma$-field in $X$ for which these mappings are measurable. Further, assume that $(X, T)$ is Hausdorff. Let $\{\mu_\alpha\}$ be a family of probability measures on $(X, T)$ and $I : X \to [0, \infty)$ a rate function with $T$-compact level sets. Then $\{\mu_\alpha\}$ satisfies the LDP with rate function $I$ if and only if for each integer $d \geq 1$ and $\alpha_1, \ldots, \alpha_d \in A$, the family of probability measures $\{\mu_\alpha \circ (\Lambda_{\alpha_1} \times \cdots \times \Lambda_{\alpha_d})^{-1}\}$ on $Y^d$ satisfies the LDP with rate function

$$I_{\alpha_1, \ldots, \alpha_d}(y_1, \ldots, y_d) = \inf \{I(x) : x \in X, \Lambda_{\alpha_i}(x) = y_i, 1 \leq i \leq d\}.$$  

Proof. Necessity. For $d \geq 1$ and $\alpha_1, \ldots, \alpha_d \in A$, the map $(\Lambda_{\alpha_1}, \ldots, \Lambda_{\alpha_d})$ from $X$ to $Y^d$ is continuous with respect to the corresponding topologies. Thus necessity follows via Lemma 2.2(a).

Sufficiency. Note that the collection

$$\{(\Lambda_{\alpha_1}, \ldots, \Lambda_{\alpha_d})^{-1}(U) : d \geq 1, \alpha_1, \ldots, \alpha_d \in A, \text{ open } U \subseteq Y^d\}$$

is a base for the topology $T$. We first show the lower bound, for $B \in \Sigma$,

$$\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B) \geq -I(\text{int}(B)).$$

For each $u \in B$, there exist $d \geq 1$, $\alpha_1, \ldots, \alpha_d \in A$, and an open set $U \subseteq Y^d$ such that $u \in (\Lambda_{\alpha_1}, \ldots, \Lambda_{\alpha_d})^{-1}(U) \subseteq \text{int}(B)$. Hence

$$\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B) \geq \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon((\Lambda_{\alpha_1}, \ldots, \Lambda_{\alpha_d})^{-1}(U)) \geq -\inf\{I(x) : (\Lambda_{\alpha_1}(x), \ldots, \Lambda_{\alpha_d}(x)) \in U\} \geq -I(u).$$

For the upper bound we show for $B \in \Sigma$ and for any $\alpha$, $0 < \alpha < I(\text{cl}(B))$, $\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B) \leq -\alpha$. Let $\Phi_\alpha = \{x \in X : I(x) \leq \alpha\}$. Since $\Phi_\alpha$ is compact and $\Phi_\alpha \cap \text{cl}(B) = \emptyset$, there exists a finite cover of $\Phi_\alpha$ of the form $\{(\Lambda_{\alpha_{i_1}}, \ldots, \Lambda_{\alpha_{i_d}})^{-1}(U_{i_1}), \ldots, (\Lambda_{\alpha_{k_1}}, \ldots, \Lambda_{\alpha_{k_d}})^{-1}(U_{k})\}$, where $U_i$ is an open set of $Y^{d_i}$ ($1 \leq i \leq k$), such that $(\Lambda_{\alpha_{i_1}}, \ldots, \Lambda_{\alpha_{i_d}})^{-1}(U_i) \cap \text{cl}(B) = \emptyset$, $1 \leq i \leq k$. Therefore,

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B) \leq \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon((\Lambda_{\alpha_{i_1}}, \ldots, \Lambda_{\alpha_{k_d}})^{-1}(U_{i_1} \times \cdots \times U_{k})) \leq -\inf\{I(x) : (\Lambda_{\alpha_{i_1}}(x), \ldots, \Lambda_{\alpha_{i_d}}(x)) \in U_{i_1} \times \cdots \times U_{k}^c, 1 \leq i \leq k\} \leq -\alpha. \square$$

The next result makes it possible to strengthen the claim of Lemma 2.3 from the weak topology to the stronger $\tau$ topology.
Lemma 3.2 Let \((S, \mathcal{S})\) be an arbitrary measurable space. Then the function \(J^{(m)}_P\) defined in Lemma 2.3 has \(\tau\)-compact level sets.

Proof. We will show that given \(0 \leq \alpha < \infty\), \(\Gamma_1 := \{\mu \in \mathcal{P}(\mathcal{S}^{(m)}) : J^{(m)}_P(\mu) \leq \alpha\}\) is sequentially compact in the \(\tau\) topology. Note that

\[
\Gamma_1 = \{Q^m \in \mathcal{P}(\mathcal{S}^{(m)}) : K(Q, P) \leq \alpha\} = \{Q^m \in \mathcal{P}(\mathcal{S}^{(m)}) : K(Q^m, P^m) \leq m\alpha\}.
\]

Define \(\Gamma = \{\mu \in \mathcal{P}(\mathcal{S}^{(m)}) : K(\mu, P^m) \leq m\alpha\}\). Then \(\Gamma\) is compact and sequentially compact in the \(\tau\) topology. For any sequence \(Q^m_n \in \Gamma_1 \subseteq \Gamma\), there exists a subsequence \(Q^m_{n_k}\) (say) which converges to \(\mu \in \Gamma\) in the topology \(\tau\). Now it remains to show that \(\mu \in \Gamma_1\). For every measurable set \(B\), we have \(Q^m_{n_k}(B^m) = (Q^m_{n_k}(B))^m \to \mu(B^m)\), i.e., \(Q^m_{n_k}(B)\) converges for every measurable set \(B\) as \(k \to \infty\). By the Vitali-Hahn-Saks theorem [23, Thm. 4.3.11], there exists a probability measure \(Q_0 \in \mathcal{P}(\mathcal{S})\) (say) such that \(Q^m_{n_k}\) converges to \(Q_0\) in the topology \(\tau\). Since for each choice of measurable sets \(B_1, \ldots, B_m\),

\[
Q^m_{n_k}(B_1 \times \cdots \times B_m) \to \mu(B_1 \times \cdots \times B_m)
\]

and

\[
Q^m_{n_k}(B_1 \times \cdots \times B_m) = \prod_{i=1}^m Q^m_{n_k}(B_i) \to Q^m_0(B_1 \times \cdots \times B_m),
\]

it follows that \(\mu = Q^m_0 \in \Gamma_1\), which implies that \(\Gamma_1\) is indeed a sequentially compact and compact set. \(\square\)

3.1 LDP’s for \(U\)- and \(V\)-empirical df’s

We now present LDP’s in the \(\tau\) topology for the \(U\)- and \(V\)-empiricals, \(\hat{G}_n\) and \(\bar{G}_n\), and for the kernel-type \(U\)- and \(V\)-empiricals, \(\hat{H}_n\) and \(\bar{H}_n\).

Theorem 3.1 Let \((S, d)\) be a separable metric space and \(\{X_1\}\) an i.i.d. \(S\)-valued sequence having probability measure \(P\).

(i) The \(U\)- and \(V\)-empirical measures \(\hat{G}_n\) and \(\bar{G}_n\) each satisfy the LDP on \(\mathcal{P}(\mathcal{S}^{(m)})\) in the \(\tau\) topology (i.e., considered as \(\mathcal{B}_\tau\)-measurable functions), with rate function \(J^{(m)}_P\) defined in Lemma 2.3.

(ii) Let \(h\) be a kernel mapping \((S^m, \mathcal{S}^{(m)})\) to an arbitrary measurable space \((Y, \mathcal{Y})\). Then the kernel-type \(U\)- and \(V\)-empirical measures \(\hat{H}_n\) and \(\bar{H}_n\) each satisfy the LDP on \(\mathcal{P}(\mathcal{Y})\) in the \(\tau\) topology, with rate function

\[
J^{(m)}_{P,h}(\nu) = \inf\{J^{(m)}_P(\mu) : \mu \in \mathcal{P}(\mathcal{S}^{(m)}), \nu = \lambda_h(\mu)\}
\]

\[
= J^{(m)}_P(\lambda_h^{-1}(\nu))
\]

\[
= \inf\{K(Q, P) : \nu = Q^m \circ h^{-1}, Q^m \in \mathcal{P}(\mathcal{S}^{(m)}), \nu \in \mathcal{P}(\mathcal{Y})\},
\]

where \(\lambda_h\) is the mapping of \(\mathcal{P}(\mathcal{S}^{(m)})\) to \(\mathcal{P}(\mathcal{Y})\) defined by \(Q^m \to Q^m \circ h^{-1}\).
PROOF. (i) The $\tau$ topology on $\mathcal{P}(\mathcal{S}^{(m)})$ is generated by the family of maps $\Lambda_f : \mathcal{P}(\mathcal{S}^{(m)}) \to \mathbb{R}$, $f \in B(S^m)$, defined by $\Lambda_f(\mu) = \int f \, d\mu$, $\mu \in \mathcal{P}(\mathcal{S}^{(m)})$. Let $\{\mu_n\}$ and $\{\tilde{\mu}_n\}$ denote the probability measures (on $\mathcal{P}(\mathcal{S}^{(m)})$) of $\{\hat{G}_n\}$ and $\{\tilde{G}_n\}$, respectively. Then, for each $f \in B(S^m)$, $\{\mu_n \circ \Lambda_f^{-1}\}$ and $\{\tilde{\mu}_n \circ \Lambda_f^{-1}\}$ are the probability measures (on $B(\mathbb{R})$) of $\{U_n(f)\}$ and $\{V_n(f)\}$, respectively. By Theorem 2.1 and Corollary 2.1, $\{\mu_n \circ \Lambda_f^{-1}\}$ and $\{\tilde{\mu}_n \circ \Lambda_f^{-1}\}$ each satisfy the LDP with rate function

$$I_f(x) = \inf\{K(Q, P) : \int_{S^m} f \, dQ^m = x, Q^m \in \mathcal{P}(\mathcal{S}^{(m)})\}, \ x \in \mathbb{R}.$$ 

Hence, by Lemmas 3.1 and 3.2, $\{\mu_n\}$ and $\{\tilde{\mu}_n\}$ each satisfy the LDP with the rate function $J_P^{(m)}$ defined in Lemma 2.3.

(ii) Let $\{\nu_n\}$ and $\{\tilde{\nu}_n\}$ denote the probability measures (on $\mathcal{P}(\mathcal{Y})$) of $\{\hat{H}_n\}$ and $\{\tilde{H}_n\}$, respectively. Since $\hat{H}_n = \tilde{G}_n \circ h^{-1}$ and $\tilde{H}_n = \tilde{G}_n \circ h^{-1}$, we have

$$\nu_n(A) = P(\hat{H}_n \in A) = P(\tilde{G}_n \circ h^{-1} \in A) = P(\tilde{G}_n \in B) = \mu_n(B),$$

where $B \subset \mathcal{P}(\mathcal{S}^{(m)})$ is such that $A = \{Q^m \circ h^{-1} : Q^m \in B\}$. Thus, for $\lambda_h$ as defined in the statement of the theorem, $\nu_n = \mu_n \circ \lambda_h^{-1}$ and $\tilde{\nu}_n = \tilde{\mu}_n \circ \lambda_h^{-1}$. Since $\lambda_h$ is a continuous mapping with respect to the $\tau$ topologies on $\mathcal{P}(\mathcal{S}^{(m)})$ and $\mathcal{P}(\mathcal{Y})$, (ii) follows by Lemma 2.2(a). \qed

Remarks 3.1 (i) It is of interest to note that Theorem 3.1(i) implies an LDP for the classical empirical measure $L_n$ in a topology even stronger than the $\tau$ topology (see Corollary 3.1 below).

(ii) In the special case of the kernel $h(x) = x$, Theorem 3.1(ii) reduces to the classical Sanov theorem in the $\tau$ topology for i.i.d. samples from a separable metric space (cf. Groeneboom et al. (1979)). In fact, Groeneboom et al. consider (for $h(x) = x$) the more general case of i.i.d. samples from a Hausdorff space. Our proof extending to the case of general kernels $h$ is, however, based on Theorem 2.1 and Corollary 2.1 and hence necessarily restricted to separable metric spaces.

(iii) Theorem 3.1(i) can be extended to multisample $U$-empirical measures in a straightforward way. \qed

3.2 New results for the classical empirical df

We define for $m \geq 1$ the $\tau_m$ topology on $\mathcal{P}(\mathcal{S})$ to be the smallest topology such that for each $f \in B(S^m)$, the map $Q \to \int f \, dQ^m$ is continuous. Clearly, all $\tau_m$ topologies are coarser than the topology of the total variation metric on $\mathcal{P}(\mathcal{S})$. It is known that the $\tau_m$ topologies form a sequence of strictly finer topologies (see [9, Exercise 7.3.18]).

Corollary 3.1 Let $(S, d)$ be a separable metric space and $\{X_i\}$ an i.i.d. $S$-valued sequence having probability measure $P$. For each $m \geq 1$, the classical empirical measure $L_n$ satisfies the LDP in the $\tau_m$ topology with rate function $J_P(Q) = K(Q, P)$, $Q \in \mathcal{P}(\mathcal{S})$.

PROOF. Since the map $Q \to Q^m$ is a homeomorphism with respect to the $\tau_m$ topology on $\mathcal{P}(\mathcal{S})$ and the $\tau$-topology on $\{Q^m : Q \in \mathcal{P}(\mathcal{S})\}$, by Lemma 3.2 the function $J_P(Q) = K(Q, P)$ has $\tau_m$-compact level sets. Therefore, the use of Lemma 3.1 and Corollary 2.1 together yields the result. \qed

In light of Corollary 3.1, it is natural to ask whether $L_n$ satisfies the LDP in the “limit topology” of the $\tau_m$ topologies. Specifically, we define the $\tau_\infty$ topology to be the smallest topology on $\mathcal{P}(\mathcal{S})$ such that for each $m \geq 1$ and $f \in B(S^m)$, the map $Q \to \int f \, dQ^m$ is continuous. We then formulate an open question.
Conjecture 3.1 The function $J_P(Q) = K(Q, P)$ has $\tau_\infty$-compact level sets, and the classical empirical measure $L_n$ satisfies the LDP in the $\tau_\infty$ topology with rate function $J_P(Q)$.

3.3 LDP’s for statistical functionals

Statistical applications of Theorem 3.1 are carried out by writing the statistics of interest as “statistical functionals,” i.e., as functionals $T(\cdot)$ which are defined on df’s and are evaluated at some notion of empirical df. Let $\mathcal{P}$ denote a class of probability measures on a measurable space $(X, \mathcal{X})$ and consider a functional $T : \mathcal{P} \to \bar{\mathbb{R}} = [\infty, \infty]$. For $t \in \mathbb{R}$, put $\Omega_t = \{\mu \in \mathcal{P} : T(\mu) \geq t\}$. In the following result, $\{F_n\}$ stands for any of the sequences $\{L_n\}, \{\hat{G}_n\}, \{\hat{G}_n\}, \{\hat{H}_n\}, \{\hat{H}_n\}$, among other possibilities.

**Theorem 3.2** Let $T$ be a $\tau$-continuous (i.e., continuous in the $\tau$ topology) functional mapping $\mathcal{P} \to \bar{\mathbb{R}}$. Let $\{F_n\}$ be a sequence of $\mathcal{B}_\tau$-measurable empirical df’s based on a sequence $\{X_n\}$ of r.v.’s having probability measure $P$. Suppose that $\{F_n\}$ satisfies the LDP on $\mathcal{P}$ in the $\tau$ topology with rate function $I_P$ and, further, that $\lambda_P(t) = I_P(\Omega_t)$ is continuous from the right at $t = r$. Then, for any sequence $\{a_n\}$ satisfying $a_n \to 0$,

$$\lim_{n \to \infty} n^{-1}\log P(T(F_n) \geq r + a_n) = -\lambda_P(r).$$

The proof of Theorem 3.2 makes use of the following result.

**Lemma 3.3** Let $T$ be a statistical functional which is upper semicontinuous in the $\tau$ topology. Let $\alpha(\mu)$ have $\tau$-compact level sets in $\mathcal{P}$. Then the function $\beta(t) = \inf_{\mu \in \Omega_t} \alpha(\mu)$ is continuous from the left.

**Proof.** The proof essentially follows that of Lemma 3.3 in Groeneboom et al. (1979). Fix $t \in \mathbb{R}$. First suppose $\beta(t) < \infty$. Since $\alpha$ has $\tau$-compact level sets and $\Omega_t$ is $\tau$-closed by upper semicontinuity, there exists $\mu_t \in \Omega_t$ such that

$$\beta(t) = \inf_{\mu \in \Omega_t} \alpha(\mu) = \alpha(\mu_t) < \infty.$$ 

Since the function $\beta$ is increasing, for any sequence $t_n \uparrow t$, the limit $\lim_{n \to \infty} \beta(t_n)$ exists and $\lim_{n \to \infty} \beta(t_n) \leq \beta(t) < \infty$. For each $n \geq 1$, there exists $\mu_n \in \Omega_t$ such that $\beta(t_n) = \inf_{\mu \in \Omega_t} \alpha(\mu) = \alpha(\mu_n) \leq \alpha(\mu_t)$. Let $\Gamma = \{\mu : \alpha(\mu) \leq \beta(t)\}$. Then clearly $\mu_n \in \Gamma$ for all $n \geq 1$. By the sequential compactness of $\Gamma$ in the $\tau$ topology, there exists a subsequence $\{\mu_{n_k}\}$ and $\mu_0 \in \Gamma$ such that $\mu_{n_k} \to \mu_0$ in the $\tau$ topology, which implies that

$$\lim_{n \to \infty} \alpha(\mu_n) = \lim_{n \to \infty} \beta(t_n) = \lim_{k \to \infty} \beta(t_{n_k}) = \lim_{k \to \infty} \alpha(\mu_{n_k}) \geq \alpha(\mu_0),$$

where the inequality follows from $\tau$-lower semicontinuity of $\alpha$. Moreover, since $T$ is $\tau$-upper semicontinuous, $\mu_{n_k} \in \Omega_{t_{n_k}}$ implies

$$T(\mu_0) \geq \lim_{k \to \infty} \sup T(\mu_{n_k}) \geq \lim_{k \to \infty} \sup t_{n_k} = t,$$

i.e., $\mu_0 \in \Omega_t$. Therefore,

$$\beta(t) \leq \alpha(\mu_0) \leq \lim_{n \to \infty} \beta(t_n) \leq \beta(t).$$
Finally, assume $\beta(t) = \infty$. For a sequence $t_n \uparrow t$, if $\beta(t_n) \leq M$ for all $n \geq 1$ and for some constant $M$, then the preceding argument implies that there exists $\mu_0 \in \Omega_t$ such that $\alpha(\mu_0) \leq M$, which is in contradiction to $\beta(t) = \infty$. \hfill \Box

**Proof of Theorem 3.2.** By utilizing Lemma 3.3, we can adapt with little modification the argument in the proof of Theorem 3.2 in Groeneboom et al. (1979). \hfill \Box

**Remark 3.2** Theorem 3.2 of Groeneboom et al. (1979) is a special case of the above Theorem 3.2. As remarked in their paper, the claim continues to hold if $T$ is an $\mathbb{R}^p$-valued function and $r$ and $a_n$ are vectors in $\mathbb{R}^p$, in which case $T := (T_1, \ldots, T_d) \geq r := (r_1, \ldots, r_d)$ means $T_i \geq r_i$ for all $i$. \hfill \Box

**Example 3.1** Generalized $L$-statistics. A wide class of “generalized $L$-statistics” (see [27], [20], [18], and [28]) is given by $T(\hat{H}_n)$, where $T$ has the form

$$ T(H) = \int_0^1 J(u) q\left(\sum_{j=1}^d a_j H^{-1}(p_{uj})\right) du, $$

with $q(\cdot)$ a continuous function, and, for each $j$, $p_{uj}$ strictly monotone in $u$. In the case that the support of $J$ is contained in $[\alpha, \beta]$ for some $0 < \alpha < \beta < 1$, the functional $T$ is weakly continuous (and hence $\tau$-continuous). For a proof let $H_n$ be a weakly convergent sequence with limit $H$. Then by continuity of $q$ and the strict monotonicity of each $p_{uj}$, $q\left(\sum_{j=1}^d a_j H_n^{-1}(p_{uj})\right)$ converges to $q\left(\sum_{j=1}^d a_j H^{-1}(p_{uj})\right)$ almost surely in $u$. Since the support of $J$ is contained in $[\alpha, \beta] \subset (0,1)$, the claim follows via the dominated convergence theorem.

Well-known special cases are $U$-statistics (take $q(x) = x$, $d = 1$, $a_1 = 1$, $p_{u1} = u$ and $J(u) \equiv 1$, and evaluate the resulting functional $T(H) = \int_0^1 H^{-1}(u) du$ at $H = \hat{H}_n$) and $L$-statistics (take $q(x) = x$, $d = 1$, $a_1 = 1$, $p_{u1} = u$, and evaluate the resulting functional $T(H) = \int_0^1 J(u) H^{-1}(u) du$ at $H = L_n$). Two other important special cases, which we shall treat in some detail here, are as follows.

(i) **Trimmed $U$-statistics.** Take $q(x) = x$, $d = 1$, $a_1 = 1$, $p_{u1} = u$, and

$$ J(u) = (1 - 2\alpha)^{-1} 1_{(\alpha,1-\alpha)}(u), $$

for some $\alpha \in (0,1/2)$. Then $T(\hat{H}_n)$ becomes a trimmed $U$-statistic,

$$ T(\hat{H}_n) = \frac{1}{1 - 2\alpha} \left[ \left( \frac{\lfloor an(m) \rfloor}{n(m)} + 1 \right) W_{n, \lfloor an(m) \rfloor + 1} + \sum_{j=\lfloor an(m) \rfloor + 2}^{n(m) - \lfloor an(m) \rfloor - 1} W_{n,j} + \left( \frac{\lfloor an(m) \rfloor}{n(m)} + 1 - \alpha \right) W_{n,n(m) - \lfloor an(m) \rfloor} \right], $$

where $W_{n,1} \leq \ldots \leq W_{n,n(m)}$ denote the ordered values of the kernel evaluations $h(X_{i_1}, \ldots, X_{i_m})$ that are summed up to form the usual $U$-statistics based on the kernel $h$. This is a variant of the $\alpha$-trimmed $U$-statistic,

$$ T_n = \frac{1}{n(m) - 2\lfloor an(m) \rfloor} \sum_{j=\lfloor an(m) \rfloor}^{n(m) - \lfloor an(m) \rfloor} W_{n,j}, $$

$$ 13 $$
Assume that \( E \exp(t_0|h(X_1, \ldots, X_m)|) < \infty \) for some \( t_0 > 0 \). Then “exponential closeness” of \( T(\tilde{H}_n) \) and \( T_n \) can be readily established, i.e., for any \( \delta > 0 \),

\[
\lim_{n \to \infty} \log P(|T(\tilde{H}_n) - T_n| \geq \delta) = -\infty.
\]

From this it follows that \( T(\tilde{H}_n) \) and \( T_n \) have the same LDP behavior, which is thus given by Theorem 3.2.

(ii) Spread measures. Here we assume that the underlying observations \( X_i \) are real-valued and have continuous cdf \( F \). Two nonparametric measures of spread proposed by Bickel and Lehmann (1979) are as follows. In one case, let \( T_1 \) denote the above functional \( T(\cdot) \) with \( q(x) = x^2, d = 2, a_1 = 1, a_2 = -1, p_{u1} = u, p_{u2} = 1 - u \), and \( J(u) = (1 - 2\beta)^{-1}1_{(\beta, 1 - \beta)}(u) \) for some \( \beta \in (0, 1/2) \), i.e.,

\[
T_1(H) = \frac{2}{1 - 2\beta} \int_{1/2}^{1-\beta} (H^{-1}(t) - H^{-1}(1 - t))^2 dt.
\]

The relevant spread measure is given by \( T_1(L_n) \) and the LDP for this statistic then follows by Theorem 3.2 in conjunction with the Sanov LDP for \( L_n \) (i.e., Lemma 2.1), provided that the corresponding rate function

\[
\lambda_P(t) = I_P(\Omega_t) = J_P(\Omega_t) = K(\Omega_t, F)
\]

is continuous from the right at \( t = r \) in \( \mathbb{R} \), where \( \Omega_t = \{ \text{cdf's} : T_1(H) \geq t \} \). Without loss of generality, assume \( r \geq 0 \). Since \( \Omega_t \) is weakly closed and \( H \mapsto K(H, F) \) has compact level sets, there exists \( H_\epsilon \in \Omega_t \) such that \( K(\Omega_\epsilon, F) = K(H_\epsilon, F) \). By a construction used in [16, proof of Thm. 6.1], for \( \epsilon > 0 \) we can find \( H_\epsilon \) such that \( T_1(H_\epsilon) > T_1(H_r) \) and \( K(H_\epsilon, F) \to K(H_r, F) \) as \( \epsilon \downarrow 0 \). Thus \( K(\Omega_t, F) \) is continuous from the right at \( t = r \).

For the other spread measure, let \( T_2 \) denote \( T(\cdot) \) with \( q(x) = x^2, d = 1, a_1 = 1, p_{u1} = (u + 1)/2, \) and \( J(u) = (1 - \alpha - \beta)^{-1}1_{(\alpha, 1 - \beta)}(u) \) for some \( \alpha, \beta \) satisfying \( 0 < \alpha < 1/2 < 1 - \beta < 1 \), i.e.,

\[
T_2(H) = \frac{1}{1 - \alpha - \beta} \int_\alpha^{1-\beta} \left[ H^{-1}\left(\frac{t + 1}{2}\right)\right]^2 dt.
\]

The relevant spread measure is then given by \( T_2(\tilde{H}_n) \), where \( \tilde{H}_n \) is based on the kernel \( h(x_1, x_2) = x_1 - x_2 \). The LDP for this statistic then follows by Theorem 3.2 in conjunction with the LDP for \( \tilde{H}_n \) (i.e., Theorem 3.1(ii)), provided that the corresponding rate function

\[
\lambda_F(t) = I_F(\Omega_t) = J^{(2)}_{F,h}(\Omega_t) = K(\{ G : H \in \Omega_t, H = G^2 \circ h^{-1} \}, F)
\]

is continuous from the right at \( t = r \) in \( \mathbb{R} \), where now \( \Omega_t = \{ \text{cdf's} : T_2(H) \geq t \} \). This right continuity is established by a similar approach as above.

**Example 3.2** Weighted Kolmogorov-Smirnov statistics. Let \( F \) be a fixed cdf and define the functional \( T_F(\cdot) \) by

\[
T_F(H) = \sup_{x \in \mathbb{R}} |H(x) - F(x)|\psi(F(x)),
\]

where \( \psi \) is finite, positive, and continuous on \( (0, 1) \). As noted in [17, Remark 5], for the case \( \psi \) bounded this functional is \( \tau \)-continuous (and for unbounded \( \psi \) \( \tau \)-continuous on a restricted set of cdf’s).

14
For $H$ given by the classical empirical df $L_n$ based on a sample of size $n$ from some cdf (not necessarily $F$), statistics of form $T_F(L_n)$ describe a well-known class of goodness-of-fit statistics. More generally, we can consider the class corresponding to $F$ given by the cdf $F_h$ of $h(X_1, \ldots, X_m)$ for some kernel $h$ and $H$ given by the associated kernel-type empirical cdf $\hat{H}_n$ considered above. For this broader class of goodness-of-fit statistics, weak convergence results have been developed by Silverman (1983). For this broader class of goodness-of-fit statistics, weak convergence results have been developed by Silverman (1983). For bounded $\psi$, we now have the LDP by direct application of Theorem 3.2. For the unbounded case we can extend Theorem 3.2 to require $\tau$-continuity only on a restricted set of cdf’s, as done in [17] for the case $m = 1$ and $h(x) = x$. For numerical computation of the relevant rate function $\lambda_P(r)$, see [17] or [30, §24.2] for useful details.

Dedication and Acknowledgment

This paper is dedicated to the distinguished academician F. Móricz on the occasion of his 60th birthday. The authors thank A. Dembo and J. Fill for comments on early versions of the paper.

References


